Journal Title
XX(X):2–36

DOI: 10.1177/ToBeAssigned
www.sagepub.com/

Malaria Intensity in Colombia by Regions and Populations

Alejandro Feged-Rivadeneira[1,2] and Andrs ngel[3] and Felipe Gonzlez[3] and Camilo Rivera[4]

Abstract

Determining the distribution of disease prevalence among heterogeneous populations at the national scale is fundamental for epidemiology and public health. Here, we use a combination of methods (spatial scan statistic, time series correlation, epidemic profile) to study measurable differences in malaria intensity by regions and populations of Colombia. This study explores three main questions: What are the regions of Colombia where malaria is epidemic? What are the regions and populations in Colombia where malaria is endemic? What associations exist between epidemic outbreaks between regions in Colombia?

Plasmodium falciparum is most prevalent in the Pacific Coast, some regions of the Amazon Basin, and some regions of the Magdalena Basin. *Plasmodium vivax* is the most prevalent parasite in Colombia, particularly in the Northern Amazon Basin, the Caribbean, and municipalities of Sucre, Antioquia and Cordoba. Malaria has been reported to be most common among 15-45 year old men. We find that the age-class suffering high risk of malaria infection ranges 20 to 30 with an acute peak at 25 years of age. Second, this pattern was not found to be generalizable across Colombian populations, Indigenous and Afrocolombian populations experience endemic malaria (with household transmission). Third, clusters of epidemic malaria for *Plasmodium vivax* were detected across Southern Colombia including the Amazon Basin and the Southern Pacific region. *Plasmodium falciparum*, was is epidemic in 13 of the 1,123 municipalities (1.2%). Some key locations act as bridges between epidemic and endemic regions. Finally, we generate a regional classification based on intensity and synchrony, dividing the country into epidemic areas and bridge areas.

Keywords

Spatial Statistic, Scan Statistic, Ourbreak Detection, Malaria, Likelihood Ratio, Endemicity, Intensity

INTRODUCTION

Malaria in Colombia has been studied from a variety of disciplines that describe disease patterns with dimensions such as the diversity of the vector (Rubio-Palis and Zimmerman 1997; Sáenz et al. 2001), characteristics of the parasite (Carmona-Fonseca 2004), social phenomena affecting disease transmission (Alexander et al. 2005; Banguero 1984; Pineda and Agudelo 2005; Arévalo-Herrera et al. 2015), and geological phenomena (Bouma et al. 1997; Poveda et al. 2001; Gagnon et al. 2002). Mainly, national and local contexts are well understood for a country that presents unusual diversity of environments and social backgrounds (including vast cultural diversity), which, in turn, represents different characteristics of malaria transmission. In contrast with Sub-Saharan Africa, where malaria is commonly a deadly disease affecting primarily children, Colombia is not considered particularly relevant in malarial disease studies given the relatively low mortality when compare with Sub Saharan Africa. However, malaria in Colombia presents certain characteristics that resemble those observed in Southeast Asia. Colombia was one of the first countries where resistance to chloroquine-based treatment was reported. Varied malaria intensity among segregated and diverse populations inhabiting different and unique environments make Colombia one of the few cases where malaria is endemic and where disease patterns are inconsistent from regionally, in contrast to several countries that follow a consistent pattern of infection, or whose segregated vulnerable populations do not differ in their epidemic patterns (Valero-Bernal 2006; WHO 2013). This does not mean that other countries have a homogeneous experience of malaria intensity across subpopulations or regions. However, disease distribution among Colombian populations has caused the parasite to generate resistance to treatment, unlike several other countries in the world except for South East Asia.

Malaria is a complex disease, and factors associated to disease severity and resistance have been reported, yet genetic resistance to malaria is more understood than to any other human disease (Hill 1992). However, the strong geographical association between resistance to the pathogen and disease severity remains a major challenge to assess the causality of human genetic resistance (Hill 1992). We know from evolutionary theory that two critical factors for selection must occur: 1) a population with genetic diversity has to exist for selection to be able to operate; 2) a differential in reproductive value of the trait in question for adaptation to evolve. Because African populations exhibit both genetic

[1]Department of Anthropology, Stanford University, CA, USA.
[2]TBD
[3]Deparment of Mathematics, Universidad de Los Andes, Colombia
[4]Walmartlabs, Sunnyvale, CA.

Corresponding author:
Alejandro Feged-Rivadeneira, Cra 14 No. 94-44 oficina 204, Data analytics, Aladino. Bogot, Colombia.
Email: afeged@slidedepot.co

Prepared using sagej.cls

diversity and experience severe malaria, genetic resistance to the pathogen appears to have emerged independently in different foci (Greenwood et al. 1991). However, unlike Africa, Colombia has no record of human genetic resistance to malaria. On the contrary, the parasite appears to have developed resistance to treatment. Until recently, it was unknown whether pathogen resistance was the result of selection of mutant strains under drug pressure, the spread of resistant strains, or adaptation of previously sensitive parasites (Rosario 1976). Genetic evidence suggests that resistant malaria emerged in at least 4 different geographical foci, consistent with the history of reports of resistant pathogens for *Plasmodium falciparum* in the Thailand-Cambodia border and Colombia in the 1950s, then spreading for two decades to South America, Asia and India, and then to Africa in Kenya and Tanzania in the late 1970s (Wellems and Plowe 2001). Resistant *Plasmodium vivax* was first reported in Papua-New Guinea in 1989, it is currently present in South East Asia, and suspected to occur in South America (Wellems and Plowe 2001). Studies have found resistant *Plasmodium vivax* at a rate of 11% in representative samples of all blood smears collected in two endemic geographical regions in Colombia: *Llanos orientales* (Eastern Plains) and Urabá (Soto et al. 2001), while others have found no evidence of resistant *Plasmodium vivax* forms in the Pacific Coast and the Amazon Basin (Castillo et al. 2002). However, therapeutic failure rates of *Plasmodium falciparum* (for representative samples of all blood smears collected) have been reported as high as 78% for these same regions (Castillo et al. 2002), and 67% in Antioquia (Blair-Trujillo et al. 2002). More resent assessments of malaria prevalence in endemic areas also suggest that uncomplicated malaria by low parasitemia is one of the biggest challenges for malaria control strategies (Arévalo-Herrera et al. 2015; Vallejo et al. 2015), and studies indicate that the observed differences are not attributable to human genetic traits that confer resistance (Ortega et al. 2015).

Few studies have addressed the malarial epidemiology by regions and populations to explore the role of malaria intensity in the emergence of resistant forms of the parasite. However, the role of Colombia in the global epidemiological context of malaria indicates that the country may present unique characteristics for disease transmission. Mainly, the presence, absence, and most importantly, emergence of resistant forms of the parasite in different regions suggests that isolated and distinct epidemic regions exist within the national boundaries, and such characteristics may play a distinctive role in the evolution of the parasite. Here we address the intensity of malaria by regions and human populations in Colombia, and the degree that the epidemic characteristics between regions affect each other.

One key aspect remains poorly understood about malaria dynamics in Colombia: how many different epidemic regions exist, and how do subpopulations in these regions experience malaria. During fieldwork, we interacted with local health officials who conducted malaria prevention programs at both local and national levels. Each public health official had knowledge and expertise about epidemic dynamics in their specific territorial assignment. However, a lack of systematic approaches hamper the ability to formalize such knowledge. Malaria intensity and the social aspects that condition the transmission of the parasite drive public health interventions. However, the regional designations are yet to be formalized based on analysis of malaria dynamics. Decisions

about prevention strategies, and how to target the most vulnerable populations are made based primarily on the expertise of local health officials, as indicated by the classification by Ruiz et al. (2006); Bouma et al. (1997); Poveda et al. (2001)

Here we generate a systematic classification of the malarious regions and subpopulations of Colombia, to characterize locations and subpopulations with epidemiological aspects of the parasite. Here we address three basic questions concerning/surrounding the intensity of malarial infection by ethnicity an region:

1. Is this population experiencing higher malaria intensity than other regions of the country?
2. Is the parasite persisting endemically within this population?
3. Are the epidemic characteristics of this subpopulation affecting other subpopulations?

Specifically, we examine five years of malarial case reports are examined for both malaria intensity, synchrony and segregation by ethnicity. First, we employ an outbreak detection algorithm (Kulldorf 1997) widely used (Kulldorff et al. 1998; Hjalmars et al. 1996; Burkom 2003) to identify clusters in space with outbreaks of malaria. Second, a stepwise data visualization technique is used to represent synchronous outbreaks, to examine areas that present similar time patterns of malarial epidemic. Finally, regional case reports are explored with descriptive statistics to analyze the intensity of malaria exposure by ethnicity.

BACKGROUND

From John Snow's seminal study of cholera in London, epidemiology has been a spatial discipline (Cameron and Jones 1983). Geographical disease patterns have been widely described for numerous pathogens and regions. We use three methods to analyze malaria in Colombia: disease clustering, disease visualization, and ecological analysis.

The production of good quality maps to understand and visualize risk of disease transmission is recognized as one of the fundamental tools for malaria control strategies, specifically, understanding the relationship between malaria endemicity and the health impact of malaria (Snow et al. 1996). Studies suggest that annual entomological inoculation rates (commonly computed as the product of the daily human biting rate, the sporozoite rates from the caught mosquitoes, and the days per year, 365 (Kilama et al. 2014)) in Ghana (100-1000), Kenya (10-60) and Gambia (less than 10) are associated to prevention of all cause childhood mortality rates by insecticide treated bed nets, with efficacy of 17%, 33%, and 63%, respectively (Snow et al. 1996). These results suggest that public health policies should vary according to malaria endemicity, since bed nets have been the linchpin of malaria prevention strategies since DDT was discontinued as a viable alternative. Yet, evidence suggests that there are several contexts in which bed nets are not efficient (Snow et al. 1996). In locations where malaria is intense, the use of bed nets is less efficient to prevent the burden of the disease.

Due to the scarcity of multi-sited studies across different countries, variation of the relationship between endemicity and overall health remains unknown (Snow et al. 1996).

However, within country variation of malaria has been subject of numerous studies. One study that addresses such relationship is produced by Omumbo et al. (1998), using GIS and malaria case reports to map malaria intensity in Kenya. Their results also question the use of treated bed nets in regions where malaria is intense, because in these communities, bednets are the most inefficient Omumbo et al. (1998).

Spatial descriptions of variation in malarial infection within countries has been addressed using maps of risk of contracting the disease. For example, Kleinschmidt et al. (2000) produced a more accurate visualization of risk of contracting malaria in Mali, by combining regression analysis with *"krigging"* (i.e., an interpolation method similar to smoothing fitted values) to account for local responses to environmental conditions such as weather, population and other topographic and sociological features. Using those methods, they are able to identify regions where the risk is higher than represented in traditional maps (Kleinschmidt et al. 2000). A similar approach, but based on entomological and demographic geo-coded records, is implemented with a GIS analysis to describe local risk of infection based upon proximity to breeding sites and human populations (Kitron et al. 1994). Beck et al. (1994) have implemented a variation of these risk maps by integrating remote sensing data to identify locations of high transmission based on human-vector interaction for a region in Mexico, including variation by season.

The applicability of mobile phone data to map human mobility with disease dynamics does pose some interesting caveats. First, the fraction of the population with high degree of mobility remains constant in some studies, but this does not necessarily mean that it is precisely that fraction of the population who is moving pathogens from one place to another (Candia et al. 2008). The Nükak represent one of the most endemic and vulnerable populations in terms of malaria persistence, and are highly mobile. Furthermore, multi-scale network models of human mobility suggest that local migration plays an important role in the synchronization of epidemics among subpopulations (Balcana et al. 2009), and suggests that small populations who are highly mobile play a fundamental role in the dispersal of epidemics.

Estimating the effect of migration on pathogen loads has been a growing interest of epidemiologists in the past years, and multiple methods have been implemented to address such interaction. A different data-driven approach to examine the effect of human mobility on epidemics has been the gravity model, used to evaluate measles outbreaks, both by age-classes and by urban and rural settings (Bharti et al. 2008; Ferrari et al. 2010a,b). The main finding of this approach was that population densities were the main driver of outbreak seasonality across different environments (Bharti et al. 2008; Ferrari et al. 2010a,b). Furthermore, the same group has used nighttime light imagery to estimate the effect of changing patterns of population densities on disease outbreaks (Bharti et al. 2011). Unfortunately, few comparative studies exist to determine which method is more effective under which conditions and for which diseases. However, the method of nighttime light imagery provides good estimates of mobility of populations without access to phone services, often the most vulnerable populations in terms of disease prevalence.

Although the gravity model has been mostly used for directly transmitted diseases, understanding the effect of human mobility on disease epidemics, and more generally how disease disperses over space and time, has been one of the fundamental questions in contemporary spatial epidemiology.

Two approaches have been used to analyze synchrony of disease outbreaks over space and time, controlling for seasonal and environmental variables. Spectral analysis has been used to describe the association between aggravation of asthma symptoms and temperature or atmospheric contamination levels (Bishop 1977; Cazelles et al. 2007), and the association between air pollution and mortality (Cazelles et al. 2007; Schwartz 1993). This technique has been used to study the effect of climatic variation on cholera (Pascual et al. 2000, 2002), malarial epidemics (Pascual et al. 2006), and the seasonality of sexually transmitted diseases (STDs) (Grassly et al. 2005).

Some authors suggest that these methods have limitations, because they can only be used for time-series data in which statistical proprieties do not change over time, yet, epidemic data are inherently complex and non-stationary (Cazelles et al. 2007). Furthermore, evidence suggests that epidemic data characteristics do change over time (Duncan et al. 1996; Rohani et al. 2003; Cazelles et al. 2007).

The limitations of the spectral decomposition methods led to the implementation of the second technique that is most widespread in understanding disease dynamics over space and time, coupled with climatic and environmental conditions from a non-stationary perspective: wavelets, a method used to show how time-series vary as a function of time and space (Cazelles et al. 2007, 2014).

Wavelet analysis has been used to study geographical hierarchies of measles epidemics, and the observed effect of vaccination policies over time (Grenfell et al. 2001). Associations between dengue epidemics and El Niño Southern Oscillation (ENSO) have also been documented using this method (Cazelles et al. 2005). Kreppel et al. (2014) found an association between ENSO, Indian Ocean Dipole (IOD) and plague dynamics in Madagascar, and Onozuka (2014) found similar effect of those two climatic phenomena on infectious gastroenteritis in Japan. Morris et al. (2014) documented that Buruli ulcer is affected by short and long rainfall patterns in French Guiana, as well as stochastic events such as ENSO. The relationship between ENSO and cutaneous leishmaniasis has also been documented for Costa Rica (Chaves and Pascual 2006). José and Bishop (2003) have studied the changing patterns and seasonality of Australian rotavirus epidemics comparing a multiplicity of methods including wavelet analysis, and detected seasonal biannual and quinquennial periods, yet, a three year epidemic period was also found to be dominant. Spectral analysis confirms that serotype harmonics interact in a complex, non-linear fashion, yielding an observable overall pattern beyond the isolated dynamics of each separate serotype, that is more than the sum of the parts, and inherent dynamics remain unchanged but the amplitude of disease infection is modified (José and Bishop 2003).

Spatial analysis methods have been applied in disease cluster identification. The main approaches used are: K-cluster analysis, detects global clusters based on each case point (Cuzick and Edwards 1990); the geographical machine (Openshaw et al. 1987) and the scan statistic (Kulldorf 1997), work by aggregating cases in different areas and

performing a hypothesis test based on a Bernoulli null model, with the advantageous difference for the scan statistic that it can perform multiple tests simultaneously. We present an implementation of the scan statistic in this case study. Small, isolated outbreaks of malaria in specific communities have been identified as "discrete mini epidemics", which represent disease severity by using space-time cluster identification (Snow et al. 1993). Disease risk by geographical location has also been implemented as simple logistic regressions that include altitude, and physical coordinates of each individual within a case-control study (Brooker et al. 2004). The scan statistic method has been used by Coleman et al. (2009) to identify disease clusters over space and time in a South African region. Zhang et al. (2008) have also implemented the scan statistic method in China to identify clusters and suggest public health resource optimization. Faires et al. (2014) used this method to identify clusters of *Clostridium difficile* over time in Ontario, Canada. Duczmal et al. (2015) implemented the scan statistic to study Chagas' disease in Brazil, while Occelli et al. (2014) do the same for end-stage renal disease (ESRD) in northen France. In Virginia, the increasing burden of Lyme disease was documented using spatiotemporal scan statistics (Li et al. 2014). In all cases, studies were able to identify areas with more cases than expected, highlighting in many cases the relevance of regions that did not present a comparatively higher incidence.

Globally, Rogers and Randolph (2000) have used maximum likelihood methods (i.e. a similar approach to the scan statistic) to predict areas where malaria is likely to expand as a result of climate change.

Malaria in Colombia

Banguero (1984) reports that malaria in Colombia is normally found in adult males from large households, and associated with their forestry activities. This finding has been explored and confirmed by knowledge, attitude and perception (KAP) studies about malaria in Colombia for different regions (Pineda and Agudelo 2005). Furthermore, other studies have shown that ethnicity in the Amazon Basin can explain a ratio of 33% of variation in risk for contracting the disease between Huitoto and Tikuna settlements (Alexander et al. 2005).

In the Americas, Barrera and collaborators (Barrera et al. 1999) studied the reintroduction of malaria to an area in Venezuela where it had been eradicated. They were able to document that the disease changed in epidemic characteristics in 1988 in the context of La Niña, while public efforts were unable to adapt, resulting in the inefficacy of public health measures under the new conditions and thus in an epidemic of malaria where it had previously been under control (Barrera et al. 1999).

Most geographical studies of malaria occur in response to the development of methods aimed at Entomological Inoculation Rate (EIR) measurements derived from the formulation of Macdonald (1955, 1956a,b), who designed EIR for the characteristics of the disease in Africa since local vectors have a high potential for transmission. Kleinschmidt et al. (2000) and Gemperli et al. (2004) analyze risk factors for malaria through the combination of case data, household surveys, and infant mortality data. From cases based in the African continent, they suggest that the EIR indicator is not

the best measurement of malaria prevalence (Gemperli et al. 2004; Kleinschmidt et al. 2000). Factors inversely related to malaria are the household's educational level, income, longer birth intervals, and possibly the age of the mother, consistent with previous studies (Gemperli et al. 2004; Kleinschmidt et al. 2000; Jain 1985).

Recently, a good amount of work on spatial analysis of disease has dealt with determining the influence of climate change on transmission. Siraj et al. (2014) showed that warmer temperatures shift the altitude at which malaria can persist in Antioquia, with an increase of 300 m. in the altitude accumulating 50% of the cases between 1994 and 1997. Because of the country's convoluted geography, and specially because altitudes just above the malaria transmission elevation limit are more densely populated, understanding the role of climate change is crucial to assess what areas will be in higher risk of transmission in the near future. Similarly, Ruiz et al. (2006) produced a multi-sited early warning system which accounts for climatic and vector variables to warn of conditions facilitating transmission, and Fuller et al. (2014) mapped the risk of malaria transmission for the whole country, based on knowledge of vector species and climatic variables. However, each method posses a limitation: Studies by Siraj et al. (2014) and Ruiz et al. (2006) only looked at a fraction of the whole country, albeit the most malarious regions of the country, while the participatory methods implemented by Fuller et al. (2014) only found weak correlations with malaria cases (having a better assessment of "potentially" malarious regions, rather than the current epidemic state).

METHODS

The analysis for this study was generated from case reports based on active and passive detection methods. All cases are laboratory confirmed and geocoded to the municipality level. We included data for 1,156 municipalities, that range in area from 15 to 65,674 km^2 ; total area sampled was 1,142 million km^2. For each municipality, we also analyzed ethnic membership, comprising 3,369 different populations.

Clustering

The two main objectives are to determine if malaria outbreaks exist in Colombia, and, if so, to determine their location. To address these objectives, we apply scan statistics to perform a hypothesis test in each municipality, examining whether it presents an outbreak. These approaches have been widely used in epidemiological studies Kulldorf (1997), Neill (2009), and Neill and Wong (2009).

The model to test hypothesis is mainly based on the Bernoulli model of Kulldorf (1997).

Given the data aggregated by municipality for 2007-2015, each record is assigned to the centroid of the municipality. Because the set of possible outbreaks (all possible aggregations of neighboring municipalities) is almost unlimited in terms of shape and size, so the step is to approximate this set. In this case, a grid G of $N \times N$ was overlaid onto Colombia's jurisdictional boundaries and then the set of possible outbreaks is limited to all the possible sub rectangles within the grid.

Now, under the Bernoulli model we consider a measurement μ for each rectangle $R \subseteq G$, where $\mu(R)$ corresponds to an integer and in our specific case, the number of individuals inside the given rectangle. This leads us to assume that there is a rectangle $Z \subseteq G$ such that each individual inside Z has a probability p of being infected, while the individuals outside Z have a probability q. Let n_R be the number of observed malaria cases inside R, so by assuming a Bernoulli and the following hypothesis for our unknown variables p and q:

$$H_0 : p = q \tag{1}$$
$$H_1 : p > q\,. \tag{2}$$

we have these possible distributions:

- Assuming H_0 :
$$n_R \sim Bin(\mu(R), p)\ \forall R \subseteq G$$
- Assuming H_1 :
$$n_R \sim Bin(\mu(R), p)\ \forall R \subseteq Z \quad \text{and} \quad n_R \sim Bin(\mu(R), q)\ \forall R \subseteq Z^C$$

And hence, under H_1, we have that Z is a region with potential malaria outbreak.

Lastly, the third and final step is to establish a measure of density for each subrectangle, in this case the likelihood ratio. This measurement of density has desirable properties to compare different sized rectangles (Neill and Wong 2009). Kulldorf (1997) derives the formula for likelihood ratio of a generic region. The scan statistic λ is defined as the highest density measurements for all subrectangles:

$$\lambda^* = \max_R\ \lambda(R)$$
$$\lambda(R) = p^{n_R}(1-p)^{\mu(R)-n_R} q^{n_G-n_R}(1-q)^{(\mu(G)-\mu(R))-(n_G-n_R)}$$

The local measurement $\lambda(R)$ can be interpreted as the likelihood that subrectangle R is an outbreak.

To test the hypothesis represented in equation 1 a Monte Carlo simulation was used to obtain the histogram of the statistic λ^* under the null hypothesis. Finally, it assesses the value of λ^* with the observed data. If $p > 0.05$ under the null model, H_0 is rejected and we assume an outbreak.

Synchronous epidemic visualization

The main objectives are to determine whether abnormal behaviors are related across municipalities and finding, if a relationship exists, groups of them that have a similar temporal patterns, independent of their geographical layout.

To address these matters we turn to topological data analysis. TDA and mapper explanation

Now, for our each one of the 1,156 municipality we calculated an epidemic occurrence vector defined and constructed as follows:

Given a certain municipality k, denote $s_{w,y}$ the w week of the year y, with $1 \leq w \leq 53$ and $2007 \leq y \leq 2015$. Now, let $\pi_k(s_{w,y})$ be the total malaria cases for the municipality k in the given week, and define the sample:

$$S_w^k = \{\pi_k(s_{i,j}) \mid w-2 \leq i \leq w+2, \quad y \in [2007, 2015]\}$$

This sample consists of the total cases reported for the municipality in question, among the surrounding weeks of w, for every year. If we let $\mu(S_w^k)$ and $\sigma(S_w^k)$ be the mean and standard deviation of the previous sample, then we can define:

$$epi_k(s_{w,y}) = \begin{cases} 1 & \text{if } \pi_k(s_{w,y}) > \mu(S_w^k) + 2\sigma(S_w^k) \\ 0 & \text{any other case} \end{cases}$$

This function tells us if a municipality shows an abnormal behaviour on the given period, that is, if at that given week, the total reported cases exceeds the mean value of the sample by two standard deviations.

The epidemic occurrence vector consists of binding the columns of the matrix defined as:

$$(V^k)_{i,j} = epi_k(S_{i,j})$$

so we have a vector indicating if the given municipality showed abnormal behaviour for each of the 477 weeks across the years 2007 to 2015.

To this new sample of vectors we applied TDA, selecting the cosine similarity the similarity notion for the records and the filter function as the first and second principal components of data.

RESULTS

Clustering

We found an outbreak of *Plasmodium vivax* malaria that comprised the Amazon Basin, including the departments of Amazonas, Caquetá, Meta, Guaviare, Putumayo and Nariño. Regions of Vichada, Chocó, Caldas and Antioquia also presented outbreaks, as did the region surrounding Barranquilla in the Caribbean. Singular clusters for this species where detected in parts of Putumayo. Significant outbreaks of *Plasmodium falciparum* malaria in municipalities of Chocó, Risaralda, Antioquia, Nariño and

Guaviare. Singular clusters for this species where detected in parts of Nariño. All significant outbreaks are highlighted in Figures 8-10.

Synchronous epidemic visualization

The TDA enabled us to find at least 5 groups with high overall disease intensity, defined as the number of cases with infants (age below 5 years) over the total number of cases. Two of this groups are respectively concentrated on the Pacific and Caribbean coasts, showing a geographic relation among their municipalities. The remaining three groups have their members scattered around different parts of the country, including Choco, Guania, Antioquia and Casanare.

We also identified five significant municipalities, corresponding to the central nodes in the TDA graph (figure 4a) and are reported in table 7. This municipalities are responsible for the connection among several nodes in their corresponding subgraphs (figure 5a) appearing in overlapping zones of the selected TDA filter.

Ethnicity

Figures 11-17 show the histograms of age reports of malaria by ethnicity and region. Two distinctive patterns consistently appearing throughout different regions of Colombia. First, an endemic profile risk was observed for the indigenous populations of Amazonas, Cauca, and Pacífico, mostly associated with *Plasmodium vivax*, except for Pacífico, where *Plasmodium falciparum* was consistently more prevalent across all ethnic groups in comparison with the rest of the country.

Populations with no ethnic denomination (ND) presented the characteristic signature of an occupational risk hazard* consistently though most of the regions studied: Amazonas, Caribe, Cauca, Oriente, and Pacífico.

The Afrocolombian population also presented a pattern consistent with occupational risk hazard in Amazonas, Caribe, Cauca, Noroeste, Oriente, and Pacífico. Interestingly, in Cauca the Afrocolombian population has a higher prevalence of *Plasmodium falciparum*, but this was not detected within the other two ethnic groups.

Both Caribe and Noroeste presented an unusual pattern compared to other regions, where the indigenous population showed patterns that could be both occupational and endemic. However, Afrocolombian ethnicity and the ND group presented a pattern mostly consistent with occupational risk hazard, but where endemicity may also have played a role in malarial transmission.

*In a histogram of case reports by age and sex, an occupational risk hazard has a unique and characteristic signature: one age class, typically for only one sex, presents an outstanding number of cases compared to any other age class. In the case of malaria in Colombia, we observed that men with no ethnic denomination of ages 20-25 were contracting malaria far more often than any other class. From this simple observation, we inferred the following: first, men of this age class were engaging in activities that posed a risk of contracting the disease. Second, women were not engaging in this activity, nor were men in other age classes. Third, there was no household transmission, since infected men were not infecting other members of their family once they ceased to engage in the risky activity.

DISCUSSION

Nothing of the new to methods has been included here

Malaria in Colombia was characterized by a different intensity, synchrony and segregation in each region. While there was a general pattern of risk throughout the country associated with occupational hazard, some populations experienced intense malaria exposure in endemic pockets. Understanding the interaction of such pockets is fundamental to designing malarial control strategies. Here we have produced a systematic approach that analysis of malaria under three dimensions.

Colombia experienced a generalized malaria outbreak in the Amazon region for the period studied. We found that there was little synchrony among the municipalities that composed the Amazon region, and that this outbreak was spatially connected to the Southern Pacific Coast. In the Amazon region, where there was relatively high degree of cultural diversity, indigenous populations experienced malaria in endemic patterns, contrary to the risk of the ND population for both the region and for the country.

The Cauca basin was characterized by different pattern: the Afrocolombian population experienced a segregated exposure to *Plasmodium falciparum* in a way that no other ethnic group did. The pattern observed for most of the country is not consistent across the Pacific region, where *Plasmodium falciparum* also persisted at a relatively higher prevalence than in the rest of the country in comparison to *Plasmodium vivax*. Most interestingly, the Cauca Basin region contained two different populations that lived in pockets of endemicity, while it is also synchronous with other regions, and furthermore it is part of a region where the scan statistic algorithm detected an outbreak.

Our findings have potential implications for malarial control. First, we found that malaria in Colombia did present different, isolated pockets with distinctive epidemic characteristics. The magnitude of such differences in epidemic characteristics is relevant in studying the pressure of anti-malarials upon the parasite, since the emergence of resistance has been reported in the country. We found that *Plasmodium falciparum* was particularly acute among the Afrocolombian population of the Cauca Basin and the Pacific region. However, in the Cauca Basin, it constituted an isolated outbreak, while in the Pacific, the outbreak was dispersed among both the Afrocolombian and the indigenous populations. Different parasite loads among ethnically and culturally distinct populations constitute the quintessential mechanism of selective pressures that are ideal for the evolution of parasites. The diversity of epidemic characteristics of malarial infection among the subpopulations of Colombia account for an ideal environment for parasite evolution, where plasmodia persist under different pressures of asymptomatic individuals, susceptible classes of ethnically distinct populations, and public health interventions using different anti-malarial strategies. Such diversity provides the necessary conditions, acting as isolated experiments, and then sharing "successful" results, for the emergence of resistant parasites.

Second, the patterns of endemicity observed in these populations suggested that prevention efforts should be population specific, and vary according to the epidemic characteristics exhibited by the parasite in the targeted population. Therapeutic failures

have been suggested to be correlated with high intestinal parasite loads (Blair-Trujillo et al. 2002). The effectiveness of bed nets has been reported to be low among populations that experience intense malaria exposure (Snow et al. 1996). We have identified populations that experienced malaria endemicity, where prevention efforts focused on the distribution of bed nets. Our findings, combined with previous knowledge suggest that public health interventions should integrate two aspects: 1) Diagnostic and treatment of asymptomatic malaria; and 2) Diagnostic and treatment of intestinal parasites (to reduce therapeutic failure).

Third, prevention strategies focusing on populations with endemic malaria would yield a reduction of occupational hazard malaria, since the occupational hazard is associated to visiting locations where malaria persists.

FIGURES

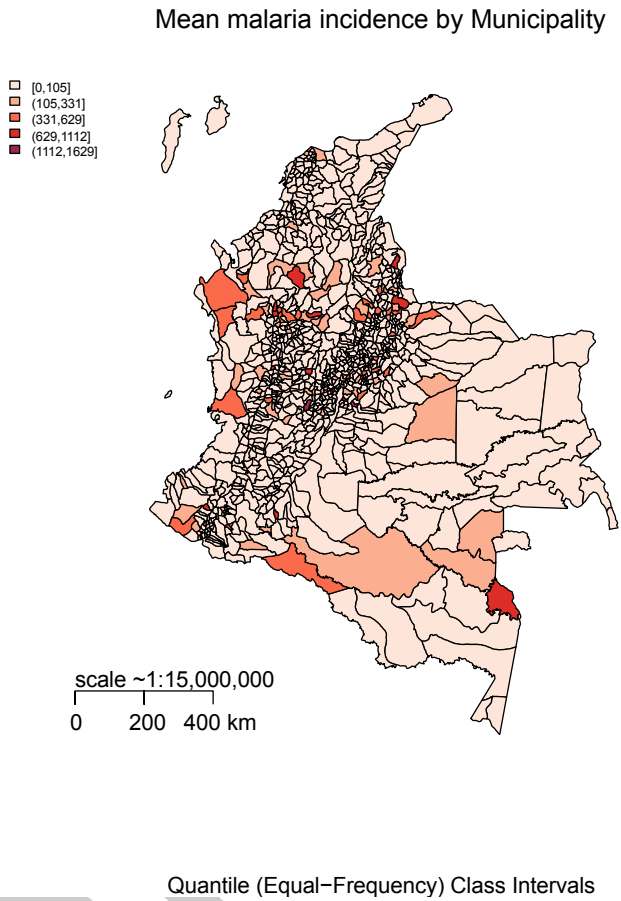


Figure 1. Malarial incidence for both species in Colombia from 2003-2008.

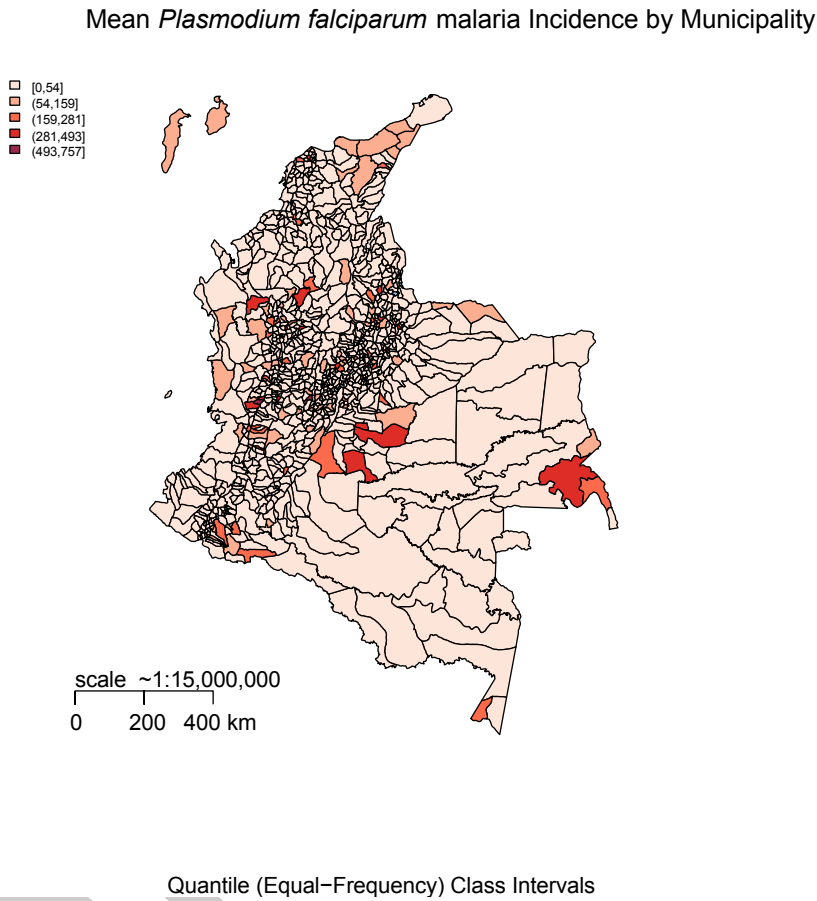


Figure 2. *Plasmodium falciparum* incidence in Colombia from 2003-2008.

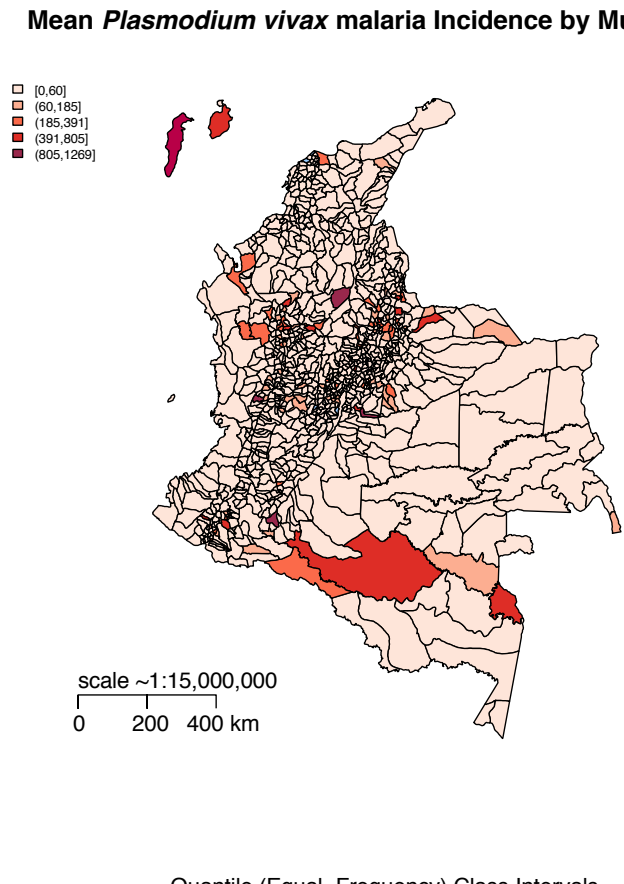
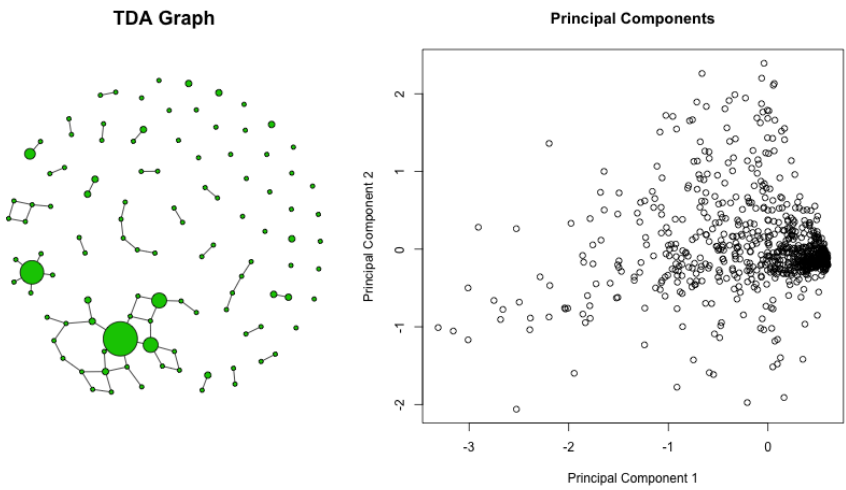


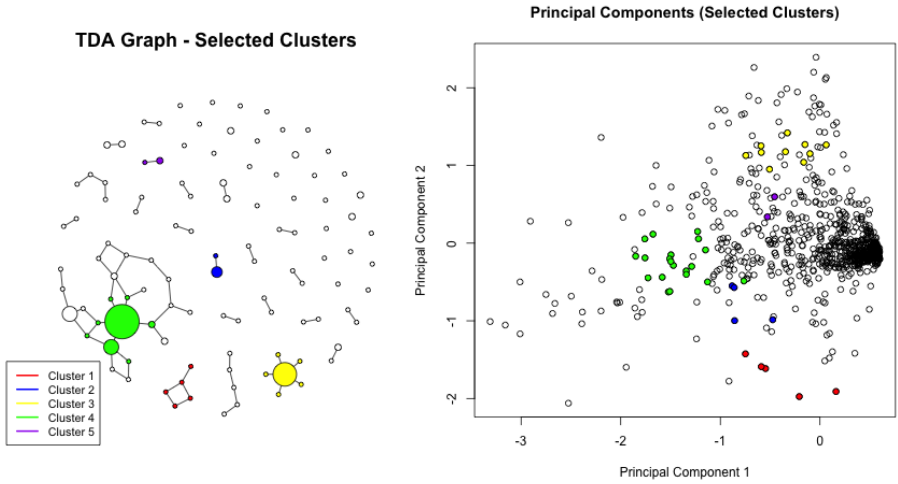
Figure 3. *P.vivax* incidence in Colombia from 2003-2008.



(a) Graph constructed using TDA over the epidemic occurrence vectors.

(b) Plot of the epidemic occurrence vectors over their first and second principal components.

Figure 4. TDA results. In the graph ach node represents a set of municipalities have similar epidemic occurrence over time, and if a municipality is in two nodes then there is an arch between them.



(a) Graph constructed using TDA over the epidemic occurrence vectors, where selected groups have been highlighted.

(b) Plot of the epidemic occurrence vectors over their first and second principal components, where the municipalities that compose the selected groups have been highlighted.

Figure 5. TDA results, with selected groups. Each of the highlighted groups represents a set of municipalities that appear in subgraphs with high overall disease intensity. So in turn, each group contains municipalities with high malaria incidence that have similar temporal behaviour.

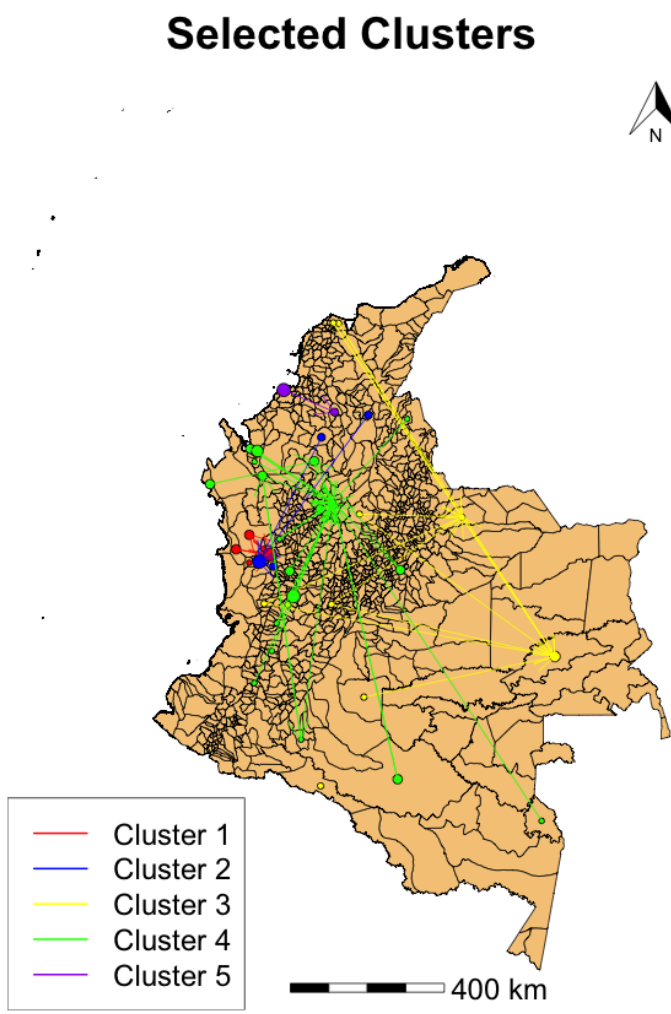


Figure 6. Selected municipalities by TDA over the Colombian territory. Notice that only the first and fifth group show a geographic pattern, whereas the rest are scattered around other territories. The arrows point to the central municipalities of each group. These central municipalities are responsible for the connectivity of their corresponding subgraphs, appearing in more than one node and in turn, creating an arch between them.

Central Nodes

Cluster	State Name	Locality	Rural Population	Urban Population	Total Population
1	Choco	Bagado	6082	2372	8454
2	Choco	Tado	6795	11246	18041
3	Guainia	Barranco Minas	4384	0	4384
4	Antioquia	Yali	4753	2981	7734
5	Sucre	Sucre	15579	6884	22463

Figure 7. Selected central municipalities after executing TDA over the epidemic occurrence vectors. This are the municipalities responsible for the connectivity among their respective groups and subgraphs.

Acknowledgements

The authors want to thank their families.

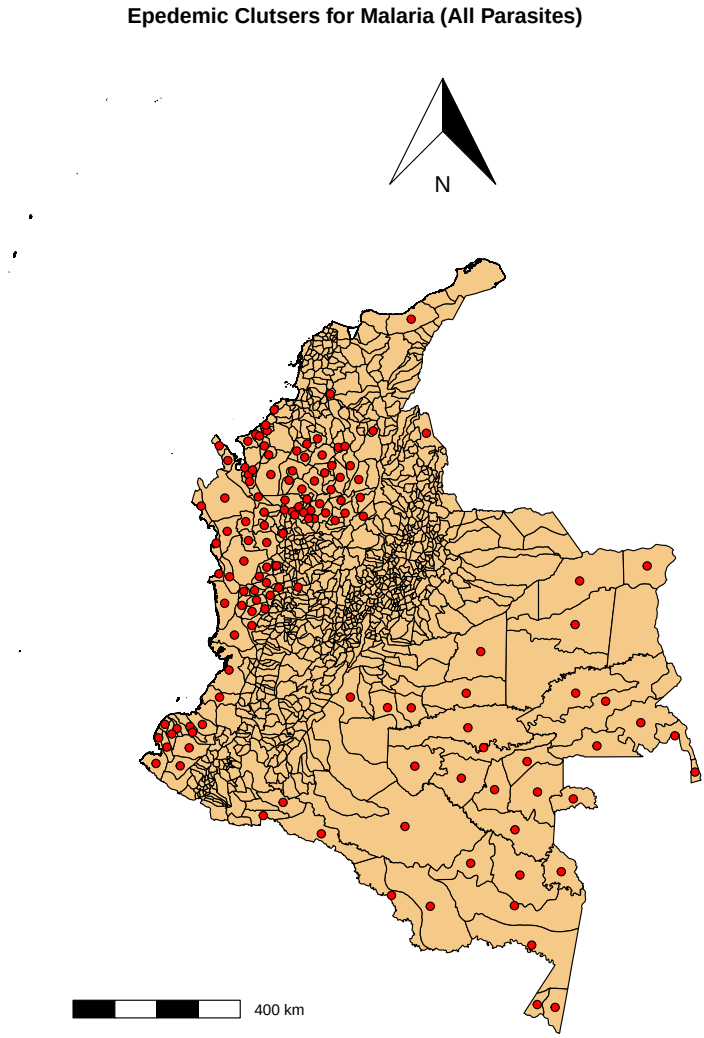


Figure 8. Significant outbreaks of malaria (all parasites) in Colombia from 2007-2015, calculated using the scan statistic developed by Kulldorf (1997) based on a likelihood ratio. The significance threshold parameter was calculated using a Bernoulli model where cases were simulated for each municipality, and taking the maximum value. The process was iterated many times and the distribution of the maximum values was calculated to determine the 95% confidence interval. The principal epidemic clusters detected by this procedure are located in Cordoba, Nariño and Antioquia

Prepared using sagej.cls

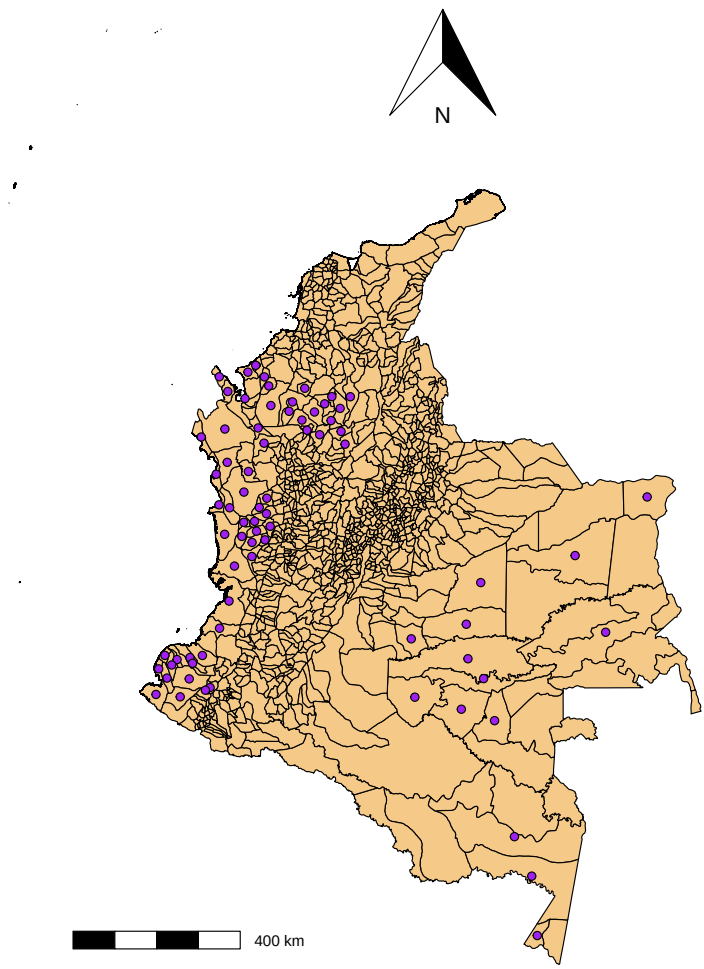


Figure 9. Significant outbreaks of *Plasmodium falciparum* in Colombia from 2007-2015. The method used to find significant outbreaks is the same as described for Figure 8. Significant clusters were observed in municipalities of departments: Chocó, Risaralda and Antioquia. The municipalities: Policarpa and Cumbitirá, in Nariño appear to be a hidden cluster for this parasite, since they weren't marked as epidemic when considering all malarian parasites. (Figure 8)

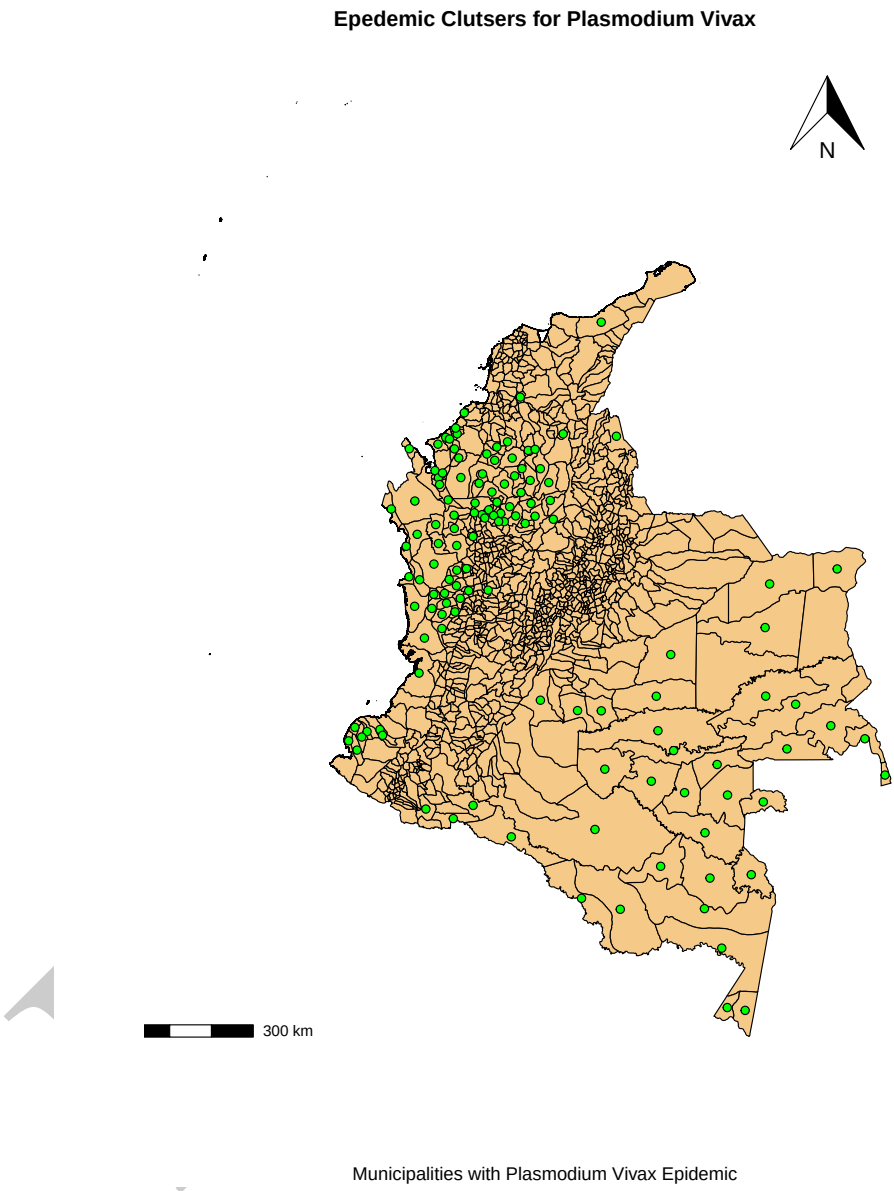


Figure 10. Significant outbreaks of *Plasmodium vivax* in Colombia from 2007-2015. The method used to find significant outbreaks is the same as described for Figure 8. Significant clusters were observed in municipalities of departments: Cordoba, Vichada and Antioquia. The municipality: Orito in Putumayo appears to be a hidden cluster for this parasite, since they weren't marked as epidemic when considering all malarian parasites. (Figure 8)

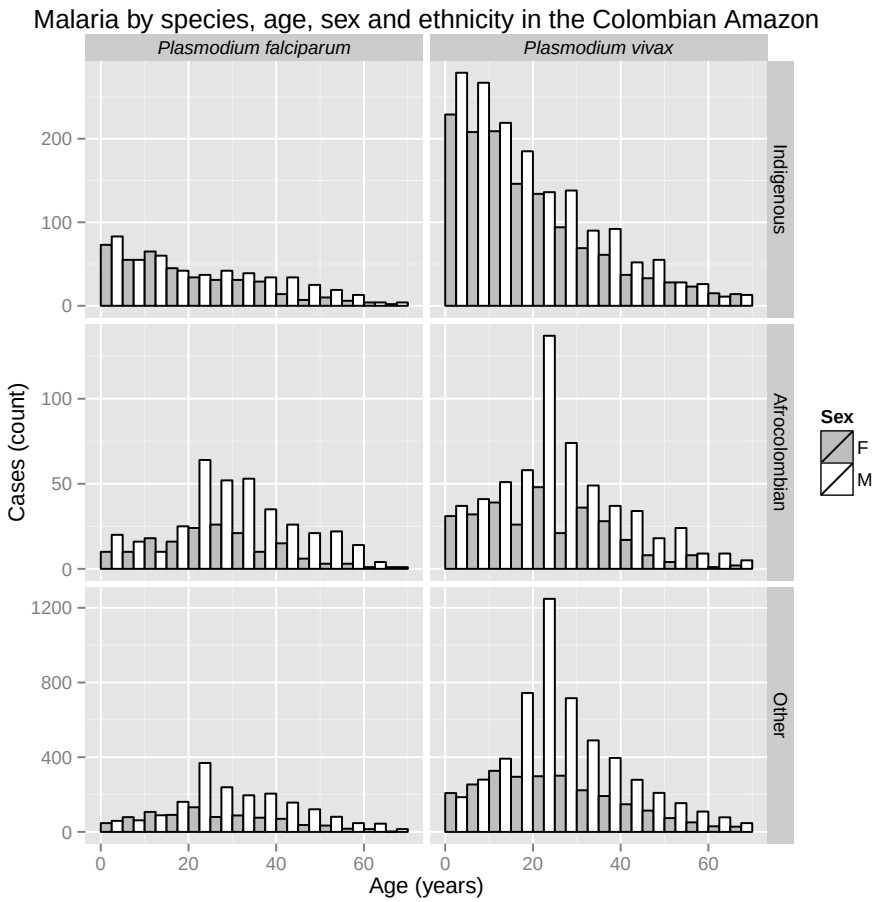


Figure 11. Malaria by parasite species, age, sex and ethnicity group of human cases in the Amazonian Colombia. Histograms show counts of cases per 5-year-age-groups. This graph shows that the indigenous population experienced endemic malaria, while the disease in the Afrocolombian and "Other" ethnic categories characterized by the profile of an occupational risk hazard.

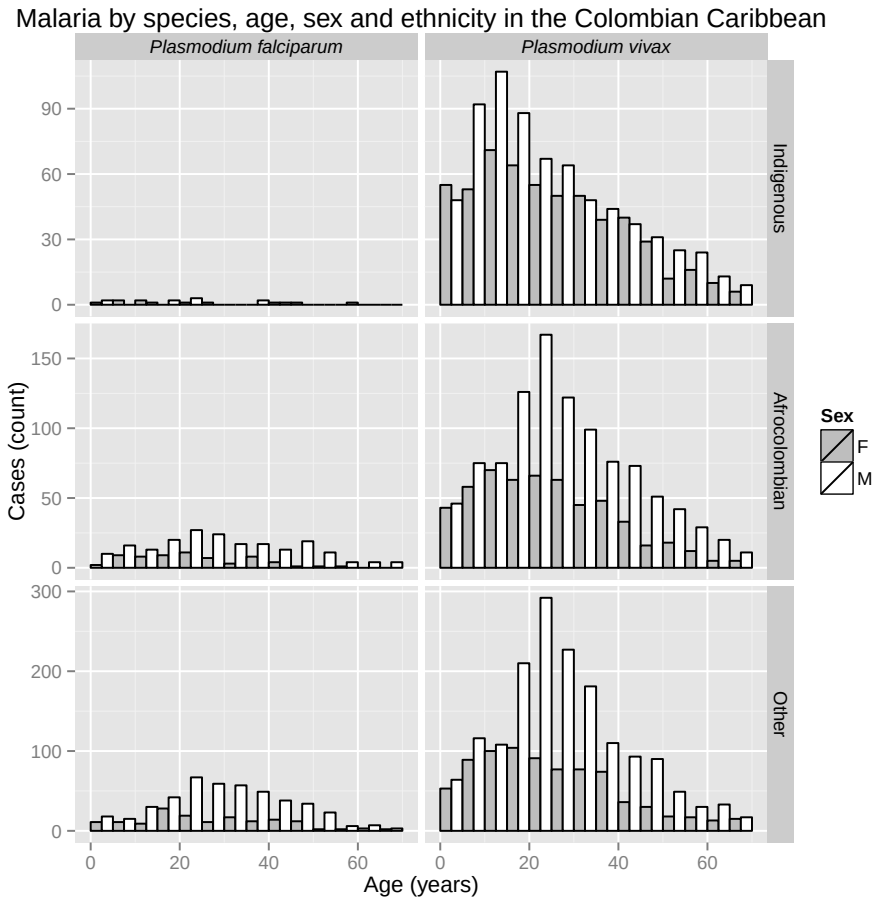


Figure 12. Malaria by parasite species, age, sex and ethnicity group of human cases in the Caribbean coast of Colombia. Histograms show counts of cases per 5-year-age-groups. This graph shows that the disease in the Afrocolombian and "Other" ethnic categories characterized by the profile of an occupational risk hazard.

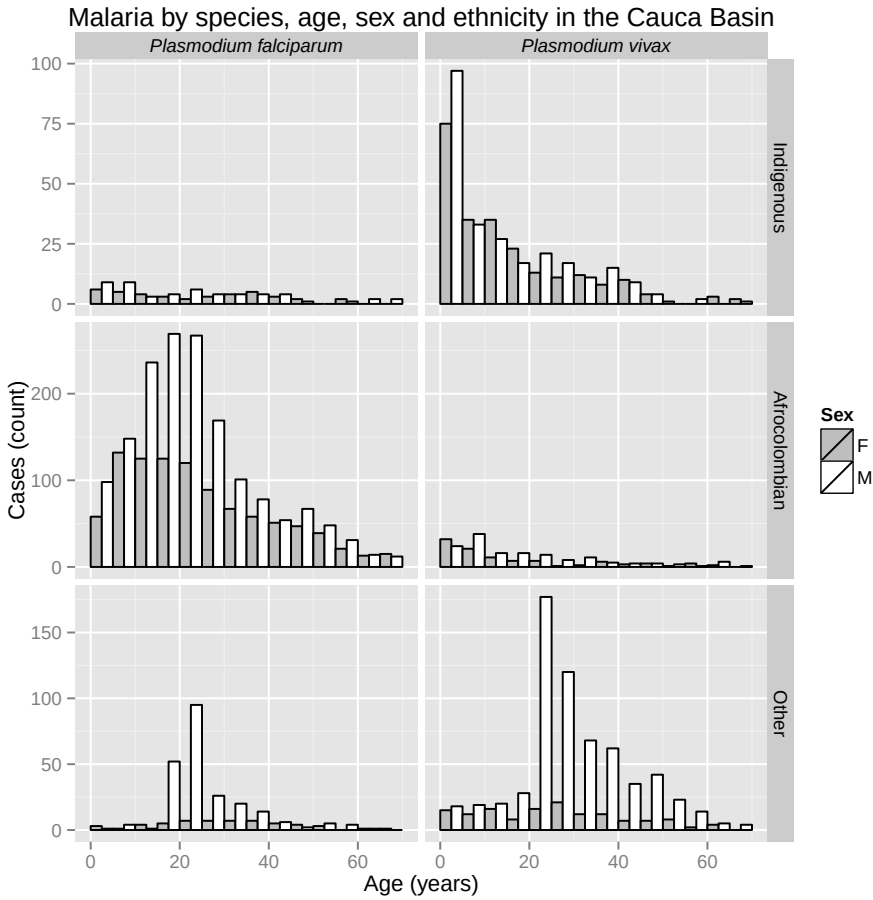


Figure 13. Malaria by parasite species, age, sex and ethnicity group of human cases in the Cauca River Basin. Histograms show counts of cases per 5-year-age-groups. This region showed a peculiar pattern compared to the rest of the country, since both the indigenous and the Afrocolombian populations showed signs of intense malaria exposure (endemic among the first and as occupational risk hazard for the second), yet, the main parasite infecting the population is different (more *Plasmodium falciparum* among Afrocolombians). This suggests that at least two separate malarious regions exist within the Cauca Basin, each with

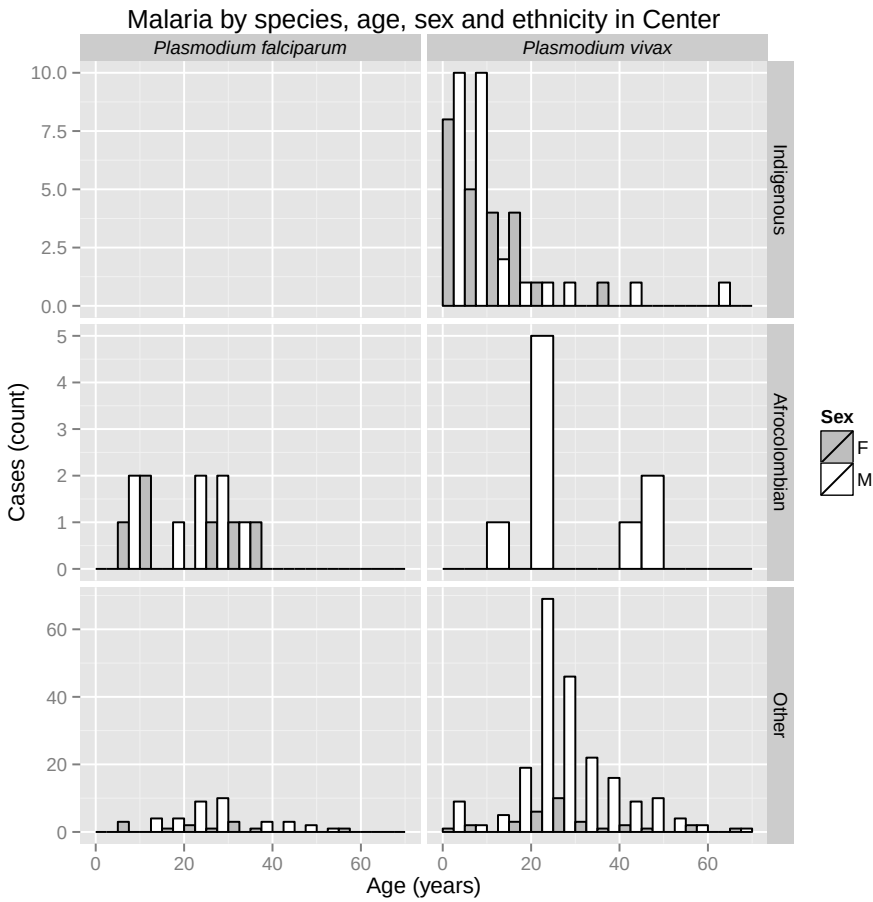


Figure 14. Malaria by parasite species, age, sex and ethnicity group of human cases in Central Colombia. Histograms show counts of cases per 5-year-age-groups. Since this is a predominantly mountainous region where there is no transmission of malaria, most cases are likely imported from elsewhere and reported at the place where they are diagnosed. An endemic pattern was observed for *Plasmodium vivax* while the "Other" category presented a pattern suggesting occupational risk hazard.

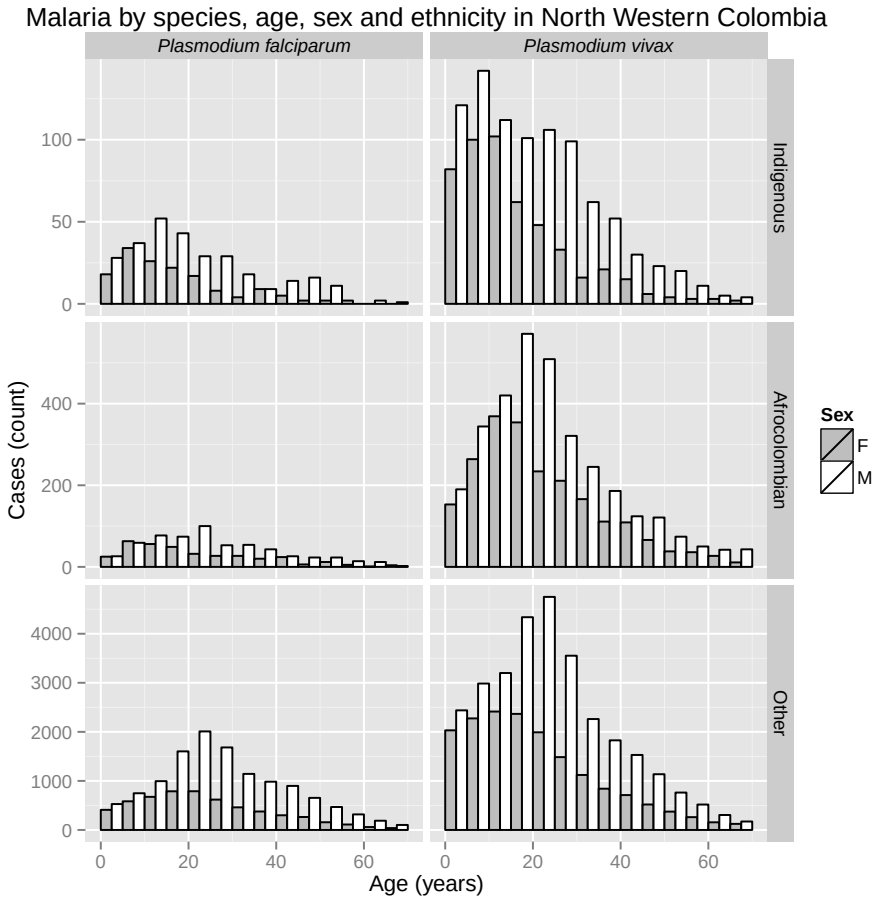


Figure 15. Malaria by parasite species, age, sex and ethnicity group of human cases in North Western Colombia. Histograms show counts of cases per 5-year-age-groups. All ethnic groups presented patterns suggesting intense malaria exposure with some degree of difference by sex. This pattern suggests that foci with both endemicity and occupational risk hazard may coexist within the region, being the indigenous population the most endemic of all.

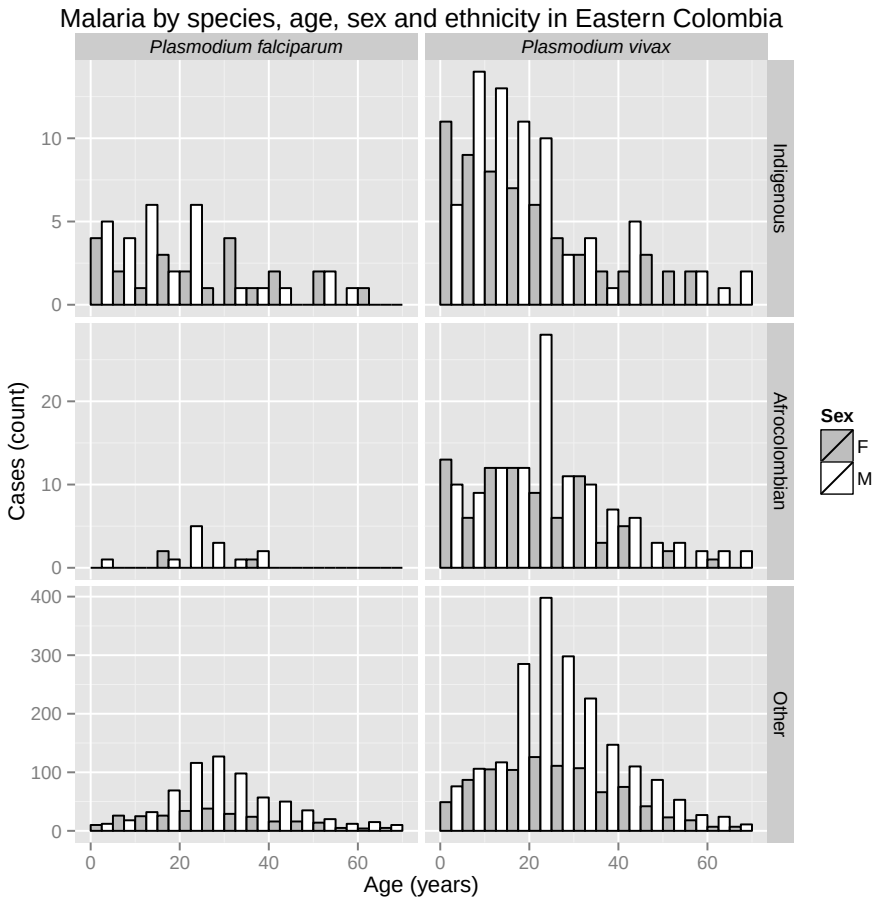


Figure 16. Malaria by parasite species, age, sex and ethnicity group of human cases in Eastern Colombia. Histograms show counts of cases per 5-year-age-groups. The malarial patterns observed suggest endemicity among the indigenous population, while a predominantly occupational risk hazard for the Afrocolombian and "Other" ethnic groups.

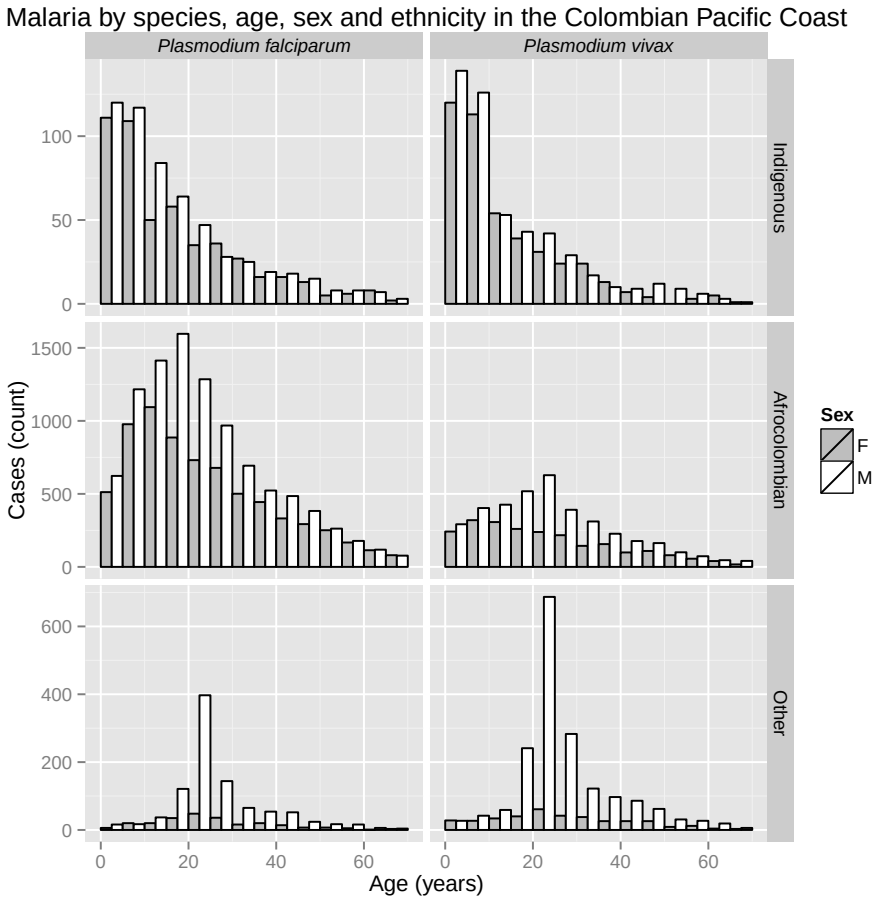


Figure 17. Malaria by parasite species, age, sex and ethnicity group of human cases in the coastal region of the Colombian Pacific. Histograms show counts of cases per 5-year-age-groups. Interestingly, *Plasmodium falciparum* was more abundant but not endemic among the Afrocolombian population, but endemic among the indigenous population. *Plasmodium vivax* was endemic among the indigenous, but not among Afrocolombians. The "Other" population group presented a pattern consistent with occupational risk hazard.

References

Yasmin Rubio-Palis and Robert H Zimmerman. Ecoregional classification of malaria vectors in the neotropics. *Journal of Medical Entomology*, 34(5):499–510, 1997.

Roberto Sáenz, Jorge A Molina, Martha Lucía Quiñones, Helena Luisa Brochero, and Víctor Alberto Olano. Mapas preliminares de la distribución de especies de anopheles vectores de malaria en Colombia. *Biomédica*, 21(4):402–408, 2001.

JAIME Carmona-Fonseca. La malaria en Colombia, Antioquia y las zonas de Urabá y Bajo Cauca: panorama para interpretar la falla terapéutica antimalárica. parte 2. *Iatreia*, 17(1):34–53, 2004.

N. Alexander, M. Rodriguez, L. Perez, J. Caicedo, C. Cruz, G. Prieto, J. Arroyo, M. Cotacio, M. Suarez, F. Hoz, and A. Hall. Case-control study of mosquito nets against malaria in the Amazon Region of Colombia. *American Journal of Tropical Medicine and Hygiene*, 73(1): 140–148, 2005.

H. Banguero. Socioeconomic factors associated with malaria in Colombia. *Social Science and Medicine*, 19(10):1099–1104, 1984. ISSN 0277-9536 (Print).

Francy Pineda and Carlos A Agudelo. Percepciones, actitudes y prácticas en malaria en el amazonas colombiano. *Revista de Salud Pública (Bogotá)*, 7(3):339–348, 2005.

Myriam Arévalo-Herrera, Mary Lopez-Perez, Luz Medina, Alberto Moreno, Juan B Gutierrez, and Sócrates Herrera. Clinical profile of *Plasmodium falciparum* and *Plasmodium vivax* infections in low and unstable malaria transmission settings of Colombia. *Malaria journal*, 14(1):154, 2015.

MJ Bouma, G Poveda, W Rojas, D Chavasse, M Quinones, J Cox, and J Patz. Predicting high-risk years for malaria in Colombia using parameters of El Niño Southern Oscillation. *Tropical Medicine & International Health*, 2(12):1122–1127, 1997.

Germán Poveda, William Rojas, Martha L Quiñones, Iván D Vélez, Ricardo I Mantilla, Daniel Ruiz, Juan S Zuluaga, and Guillermo L Rua. Coupling between annual and ENSO timescales in the malaria-climate association in Colombia. *Environmental health perspectives*, 109(5): 489, 2001.

Alexandre S Gagnon, Karen E Smoyer-Tomic, and Andrew B Bush. The El Nino Southern Oscillation and malaria epidemics in South America. *International Journal of Biometeorology*, 46(2):81–89, 2002.

María V Valero-Bernal. Malaria in Colombia: retrospective glance during the past 40 years. *Revista de Salud Pública*, 8(3):141–149, 2006.

WHO. *Malaria entomology and vector control*. World Health Organization, 2013.

Adrian VS Hill. Malaria resistance genes: a natural selection. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 86(3):225–232, 1992.

B Greenwood, K Marsh, and R Snow. Why do some African children develop severe malaria? *Parasitology Today*, 7(10):277–281, 1991.

VE Rosario. Genetics of chloroquine resistance in malaria parasites. *Nature Publishing Group*, 261(5561):585–586, 1976.

Thomas E Wellems and Christopher V Plowe. Chloroquine-resistant malaria. *Journal of Infectious Diseases*, 184(6):770–776, 2001.

J Soto, J Toledo, P Gutierrez, M Luzz, N Llinas, N Cedeño, M Dunne, and J Berman. *Plasmodium vivax* clinically resistant to chloroquine in Colombia. *The American Journal of Tropical Medicine and Hygiene*, 65(2):90–93, 2001.

Carmen Manuela Castillo, Lyda Elena Osorio, and Gloria Inés Palma. Assessment of therapeutic response of *Plasmodium vivax* and *Plasmodium falciparum* to chloroquine in a malaria transmission free area in Colombia. *Memórias do Instituto Oswaldo Cruz*, 97(4):559–562, 2002.

Silvia Blair-Trujillo, Leidy Lacharme-Lora, and Jaime Carmona-Fonseca. Resistance of *Plasmodium falciparum* to antimalarial drugs in Zaragoza (Antioquia, Colombia), 1998. *Memórias do Instituto Oswaldo Cruz*, 97(3):401–406, 2002.

Andres F Vallejo, Pablo E Chaparro, Yoldy Benavides, Álvaro Álvarez, Juan P Quintero, Julio Padilla, Myriam Arévalo-Herrera, and Sócrates Herrera. High prevalence of sub-microscopic infections in Colombia. *Malaria journal*, 14(1):201, 2015.

Diana Carolina Ortega, Cristian Fong, Heiber Cardenas, and Guillermo Barreto. Evidence of over-dominance for sickle cell trait in a population sample from Buenaventura, Colombia. *International Journal of Genetics and Molecular Biology*, 7(1):1–7, 2015.

Daniel Ruiz, Germán Poveda, Iván D Vélez, Martha L Quiñones, Guillermo L Rúa, Luz E Velásquez, and Juan S Zuluaga. Modelling entomological-climatic interactions of plasmodium falciparum malaria transmission in two colombian endemic-regions: contributions to a national malaria early warning system. *Malaria Journal*, 5(1):66, 2006.

M. Kulldorf. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 6(26): 1481–1496, 1997.

M Kulldorff, WF Athas, EJ Feurer, BA Miller, and CR Key. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health*, 88(9):1377–1380, 1998.

Ulf Hjalmars, Martin Kulldorff, GÖran Gustafsson, and Neville Nagarwalla. Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine*, 15(7-9):707–715, 1996.

Howard S Burkom. Biosurveillance applying scan statistics with multiple, disparate data sources. *Journal of Urban Health*, 80(1):i57–i65, 2003.

Donald Cameron and Ian G Jones. John Snow, the Broad Street pump and modern epidemiology. *International Journal of Epidemiology*, 12(4):393–396, 1983.

RW Snow, K Marsh, and D Le Sueur. The need for maps of transmission intensity to guide malaria control in Africa. *Parasitology Today*, 12(12):455–457, 1996.

Maxwell Kilama, David L Smith, Robert Hutchinson, Ruth Kigozi, Adoke Yeka, Geoff Lavoy, Moses R Kamya, Sarah G Staedke, Martin J Donnelly, Chris Drakeley, et al. Estimating the annual entomological inoculation rate for *Plasmodium falciparum* transmitted by *Anopheles gambiae s.l.* using three sampling methods in three sites in Uganda. *Malaria journal*, 13(1): 111, 2014.

J Omumbo, J Ouma, B Rapuoda, MH Craig, D Lesueur, and RW Snow. Mapping malaria transmission intensity using geographical information systems GIS: an example from Kenya. *Annals of Tropical Medicine and Parasitology*, 92(1):7–21, 1998.

I. Kleinschmidt, M. Bagayoko, GPY Clarke, M. Craig, and D. Le Sueur. A spatial statistical approach to malaria mapping. *International Journal of Epidemiology*, 29(2):355–361, 2000.

Uriel Kitron, Hedva Pener, Corina Costin, Laor Orshan, Zalman Greenberg, and Uri Shalom. Geographic information system in malaria surveillance: mosquito breeding and imported cases in Israel, 1992. *The American Journal of Tropical Medicine and Hygiene*, 50(5):550–556, 1994.

Louisa R Beck, Mario H Rodriguez, Sheri W Dister, Americo D Rodriguez, Eliska Rejmankova, Armando Ulloa, Rosa A Meza, Donald R Roberts, Jack F Paris, and Michael A Spanner. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *The American Journal of Tropical Medicine and Hygiene*, 51(3):271–280, 1994.

Julián Candia, Marta C González, Pu Wang, Timothy Schoenharl, Greg Madey, and Albert-László Barabási. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224015, 2008.

Duygu Balcana, Vittoria Colizzac, Bruno Gonçalvesa, Hao Hud, José J Ramascob, and Alessandro Vespignania. Multiscale mobility networks and the spatial spreading of infectious diseases. *PNAS*, 106(51), 2009.

Nita Bharti, Yingcun Xia, Ottar N Bjornstad, and Bryan T Grenfell. Measles on the edge: coastal heterogeneities and infection dynamics. *PloS One*, 3(4):e1941, 2008.

MJ Ferrari, A Djibo, RF Grais, BT Grenfell, and ON Bjørnstad. Episodic outbreak bias estimates of age-specific force of infection: a corrected method using measles as an example. *Epidemiology and Infection*, 138(01):108–116, 2010a.

Matthew J Ferrari, Ali Djibo, Rebecca F Grais, Nita Bharti, Bryan T Grenfell, and Ottar N Bjornstad. Rural–urban gradient in seasonal forcing of measles transmission in Niger. *Proceedings of the Royal Society B: Biological Sciences*, 277(1695):2775–2782, 2010b.

Nina Bharti, Andrew J Tatem, Matthew J Ferrari, Rebecca F Grais, Ali Djibo, and Bryan T Grenfell. Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science*, 334(6061):1424–1427, 2011.

Yvonne M Bishop. Statistical methods for hazards and health. *Environmental Health Perspectives*, 20:149, 1977.

Bernard Cazelles, Mario Chavez, Guillaume Constantin de Magny, Jean-Francois Guégan, and Simon Hales. Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society Interface*, 4(15):625–636, 2007.

Joel Schwartz. Air pollution and daily mortality in Birmingham, Alabama. *American Journal of Epidemiology*, 137(10):1136–1147, 1993.

Mercedes Pascual, Xavier Rodó, Stephen P Ellner, Rita Colwell, and Menno J Bouma. Cholera dynamics and El Niño-southern oscillation. *Science*, 289(5485):1766–1769, 2000.

Mercedes Pascual, Menno J Bouma, and Andrew P Dobson. Cholera and climate: revisiting the quantitative evidence. *Microbes and Infection*, 4(2):237–245, 2002.

Mercedes Pascual, JA Ahumada, LF Chaves, Xavier Rodo, and Menno Bouma. Malaria resurgence in the East African highlands: temperature trends revisited. *Proceedings of the National Academy of Sciences*, 103(15):5829–5834, 2006.

Nicholas C Grassly, Christophe Fraser, and Geoffrey P Garnett. Host immunity and synchronized epidemics of syphilis across the United States. *Nature*, 433(7024):417–421, 2005.

CJ Duncan, SR Duncan, and Susan Scott. Whooping cough epidemics in London, 1701-1812: infection dynamics, seasonal forcing and the effects of malnutrition. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1369):445–450, 1996.

P Rohani, CJ Green, NB Mantilla-Beniers, and BT Grenfell. Ecological interference between fatal diseases. *Nature*, 422(6934):885–888, 2003.

Bernard Cazelles, Kévin Cazelles, and Mario Chavez. Wavelet analysis in ecology and epidemiology: impact of statistical tests. *Journal of The Royal Society Interface*, 11(91): 20130585, 2014.

BT Grenfell, ON Bjørnstad, and J Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, 2001.

Bernard Cazelles, Mario Chavez, Anthony J McMichael, and Simon Hales. Nonstationary influence of El Niño on the synchronous dengue epidemics in Thailand. *PLoS Medicine*, 2 (4):e106, 2005.

Katharina S Kreppel, Cyril Caminade, Sandra Telfer, Minoarison Rajerison, Lila Rahalison, Andy Morse, and Matthew Baylis. A non-stationary relationship between global climate phenomena and human plague incidence in Madagascar. *PLoS Neglected Tropical Diseases*, 8(10), 2014.

Daisuke Onozuka. Effect of non-stationary climate on infectious gastroenteritis transmission in Japan. *Scientific reports*, 4, 2014.

Aaron Morris, Rodolphe E Gozlan, Hossein Hassani, Demetra Andreou, Pierre Couppié, and Jean-François Guégan. Complex temporal climate signals drive the emergence of human water-borne disease. *Emerging Microbes & Infections*, 3(8):e56, 2014.

Luis Fernando Chaves and Mercedes Pascual. Climate cycles and forecasts of cutaneous leishmaniasis, a nonstationary vector-borne disease. *PLoS medicine*, 3(8):e295, 2006.

Marco V José and Ruth F Bishop. Scaling properties and symmetrical patterns in the epidemiology of rotavirus infection. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1438):1625–1641, 2003.

Jack Cuzick and Robert Edwards. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 73–104, 1990.

Stan Openshaw, Martin Charlton, Colin Wymer, and Alan Craft. A mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information System*, 1(4):335–358, 1987.

RW Snow, JRM Armstrong Schellenberg, N Peshu, D Forster, CRJC Newton, PA Winstanley, I Mwangi, C Waruiru, PA Warn, C Newbold, et al. Periodicity and space-time clustering of severe childhood malaria on the coast of Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 87(4):386–390, 1993.

Simon Brooker, Siân Clarke, Joseph Kiambo Njagi, Sarah Polack, Benbolt Mugo, Benson Estambale, Eric Muchiri, Pascal Magnussen, and Jonathan Cox. Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. *Tropical Medicine and International Health*, 9(7):757–766, 2004.

Marlize Coleman, Michael Coleman, Aaron M Mabuza, Gerdalize Kok, Maureen Coetzee, and David N Durrheim. Using the satscan method to detect local malaria clusters for guiding malaria control programmes. *Malar J*, 8(68):10–1186, 2009.

Wenyi Zhang, Liping Wang, Liqun Fang, Jiaqi Ma, Youfu Xu, Jiafu Jiang, Fengming Hui, Jianjun Wang, Song Liang, Hong Yang, et al. Spatial analysis of malaria in Anhui province, China. *Malaria Journal*, 7(206):19, 2008.

Meredith C Faires, David L Pearl, William A Ciccotelli, Olaf Berke, Richard J Reid-Smith, and J Scott Weese. Detection of *Clostridium difficile* infection clusters, using the temporal scan statistic, in a community hospital in southern Ontario, Canada, 2006–2011. *BMC infectious diseases*, 14(1):254, 2014.

Luiz H Duczmal, Gladston J Moreira, Luis Paquete, David Menotti, Ricardo Takahashi, and Denise Burgarelli. Dry climate as a predictor of Chagas disease irregular clusters: A covariate study. *Online Journal of Public Health Informatics*, 7(1), 2015.

Florent Occelli, Annabelle Deram, Michaël Génin, Christian Noël, Damien Cuny, François Glowacki, et al. Mapping end-stage renal disease (ESRD): Spatial variations on small area level in northern France, and association with deprivation. *PLoS ONE*, 9(11), 2014.

Jie Li, Korine N Kolivras, Yili Hong, Yuanyuan Duan, Sara E Seukep, Stephen P Prisley, James B Campbell, and David N Gaines. Spatial and temporal emergence pattern of Lyme disease in Virginia. *The American Journal of Tropical Medicine and Hygiene*, 91(6):1166–1172, 2014.

David J Rogers and Sarah E Randolph. The global spread of malaria in a future, warmer world. *Science*, 289(5485):1763–1766, 2000.

R. Barrera, M.E. Grillet, Y. Rangel, J. Berti, and A. Aché. Temporal and spatial patterns of malaria reinfection in northeastern Venezuela. *The American Journal of Tropical Medicine and Hygiene*, 61(5):784, 1999.

G. Macdonald. The measurement of malaria transmission. *Proceedings of the Royal Society of Medicine*, 48(4):295–302, 1955.

G. Macdonald. Theory of the eradication of malaria. *Bulletin of the World Health Organization*, 15(3-5):369–387, 1956a.

G. Macdonald. Epidemiological basis of malaria control. *Bulletin of the World Health Organization*, 15(3-5):613–626, 1956b.

A. Gemperli, P. Vounatsou, I. Kleinschmidt, M. Bagayoko, C. Lengeler, and T. Smith. Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. *American Journal of Epidemiology*, 159(1):64, 2004.

A.K. Jain. Determinants of regional variations in infant mortality in rural India. *Population Studies*, 39(3):407–424, 1985.

AS Siraj, M Santos-Vega, MJ Bouma, D Yadeta, D Ruiz Carrascal, and M Pascual. Altitudinal changes in malaria incidence in highlands of Ethiopia and Colombia. *Science*, 343(6175): 1154–1158, 2014.

DO Fuller, A Troyo, TO Alimi, and JC Beier. Participatory risk mapping of malaria vector exposure in Northern South America using environmental and population data. *Applied Geography (Sevenoaks, England)*, 48:1–7, 2014.

D.B. Neill. Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting*, 25(3):498–517, 2009.

D. Neill and W. Wong. A tutorial on event detection. *Presented at the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2009.